

# EVALUATION CONNECTIONS

## CONTENT

A Message from the President	1
Editorial: Connecting the dots in evaluation	2
A European evaluation theory tree	3
Credible Evidence – The Challenges of Measurement	5
Rural differentiation and targeting: evaluating IFAD's approach	6
Does impact evaluation in development matter? Well, it depends what it's for!	8
Let the learning define the method rather than the method defining the learning	9
Dealing with complexity through "actor-focused" planning, monitoring & evaluation (PME)	11
Faster, smarter and cheaper: hypothesis-testing in policy and program evaluation	13
The authors	14
Evaluation News: Public involvement in research: Getting the measure of success	16

## A Message from the President

After two years at the helm of the European Evaluation Society this is my last message to Connections readers as President. I was proud to represent you at the excellent Third International Conference on National Evaluation Capacities (NEC) 2013, organized in Sao Paulo by the United Nations Development Program and the Government of Brazil (<http://www.nec2013.org/>).

For the first time, the NEC conference not only involved national governments, but also evaluation practitioners, academics and voluntary associations. It was highly participatory and allowed fulsome debate among stakeholders. It tackled three interlinked themes: Independence, Credibility and Use. In my keynote address I reached beyond the usual methodological concerns and identified other factors of evaluation credibility – the values on which the evaluation rests; the expertise of evaluators; the independence of evaluation governance; the transparency of evaluation processes; the credibility of evaluation evidence; how the evaluation is shared, communicated and reported, etc.

Specifically I described six challenges: 1) *political relevance*: evaluations should be tailored to the specific force field of the local context; 2) *stakeholder participation*: who participates and has a say matters a lot to evaluation credibility; 3) *institutional legitimacy*: credibility also hinges on the nature of the authorizing environment within which the evaluation

is conceived, planned and managed; 4) *evaluation capacity*: the knowledge, skills and dispositions not only of evaluators, but also of evaluation managers and commissioners are essential ingredients of evaluation credibility; 5) *epistemological clarity*: we should acknowledge the limitations of our measuring instruments and give careful thought as to *what counts*; and 6) *effectiveness of outreach*: high quality reporting; expert communications; ready accessibility of evaluation knowledge are critical ingredients of credibility; etc. I concluded by highlighting the importance of evaluation associations and networks in the promotion of evaluation excellence.

I also attended the Annual meeting of the American Evaluation Association in Washington DC (<http://www.eval.org/p/cm/ld/fid=21>). Given its huge size the AEA conference in Washington DC captured a wide range of evaluation ideas and concerns. I was again impressed by the enthusiasm of our American colleagues. I found it comforting to be among so many people who share a passion for the evaluation endeavour. I was also struck by the differences in perspectives and evaluative traditions between America and Europe thus confirming the wisdom of Nicoletta Stame's reflections included in the leading article of this Connections issue.

Among the numerous debates taking place in Washington this year, a salient one attracted my attention and that of many other

AEA delegates. It concerned “evaluative thinking”. For Michael Scriven we need to inject evaluative thinking in the social sciences and upgrade the role of values in the evaluative process. For Thomas Schwandt evaluative thinking is critical thinking: it should be reflective and systematic in the analysis of

evidence and the values that underlie evaluative arguments. For Sharon Rallis it includes moral reasoning. For Robin Miller it requires domain specific-knowledge. I look forward to further exploration of this topic with our American colleagues in Dublin. We expect a record attendance from all corners of

the world. I hope that you will join us too. In order to make the 11<sup>th</sup> Biennial Conference unforgettable the entire EES board would be pleased to receive your ideas and contributions.

Maria Bustelo, EES President

## EDITORIAL: CONNECTING THE DOTS IN EVALUATION

Robert Picciotto

As individuals and citizens we evaluate all the time. But we do not always do a good job of it. Daniel Kahneman secured a Nobel Prize in economics by demonstrating *inter alia* that human beings do not act according to the rational choice model dear to economists. He proved that the “remembering self” gives disproportionate weight to most recent perceptions. His experiments demonstrated that retrospective as well as forward looking assessments are subject to systematic biases.

Passions and interests cloud judgment. Throwing good money after bad is widespread. Investors tend not to cut their losses because doing so would be to admit failure. Organizations as well as individuals are creatures of habit. Path dependence in decision making is common. Successful organizations are especially at risk since they tend to ignore changes in the operating environment and as a result underperform because they have failed to adjust their objectives and approaches.

Kahneman’s findings have opened up a large new field of study (*behavioural economics*). It concentrates on how to get better public outcomes through ‘nudges’ designed to help people make decisions that are in the public interest.

Similarly evaluators seek to redress irregularities in behaviour through systematic fact finding and analysis, e.g. assessment of social programs involves theories about how people will react to incentives or how effectively they will use social services.

Double and triple loop learning are meant to help manage design and implementation

risks. This is ultimately why evaluation as a discipline fulfils a useful role in society. In a nutshell evaluators look for evidence about how things work in ways that are different from those we use in our daily decision making. Well conducted evaluations carried out independently help to overcome systemic biases.

Evaluators withhold judgment until they have identified the right questions, discovered the right ways to tackle them and collected and analysed relevant evidence. This expert and systematic process and its outcome are what we call evaluation. It has brought forth a wealth of evaluation approaches and models. This is what Alkin and Christie’s metaphorical evaluation theory tree illustrates.

In the leading article of this Newsletter Nicoletta Stame is in pursuit of a European equivalent to the original version that displayed the names of eminent American evaluation thinkers on three distinctive clusters of disciplinary branches connoting values, methods and use.

Stame’s article identifies several European individuals who deserve a place in evaluation history since they have made major contributions to the progress of the evaluation discipline. But rather emphasizing the intellectual prowess of individuals she views evaluation research on this side of the Atlantic as a collective, multi-faceted endeavour shaped by Europe’s institutions and cultures.

Inevitably the evaluation tree metaphor spawns more questions than it answers. One such question is explored in Ann Doucette’s article: it probes the neglected interface be-

tween evaluation models and the measurement instruments used to gather and assess evidence. Measurement methods influence evaluation findings more than usually acknowledged.

For example averages of collected data may be misleading if the intensity of responses or the reliability of sources is not taken into account. Equally the chosen bandwidth of observations is often too limited and as a result it may fail to capture the full range of relevant knowledge. What gets measured is as important as how accurate the measurement is. This is the thrust of Mattia Prayer Galletti’s article which stresses the role of effective targeting (and its systematic tracking) in enhancing the poverty reduction impact of rural development interventions.

Use and methods are closely related. In turn striking a better balance between values and methods is the theme struck by Irene Guijt and Chris Roche. Their trenchant article suggests that impact assessments now dominated by methodological concerns would be more credible and useful if they were explicit about *whose* learning counts, *whose* accountability is improved and *whose* influence is strengthened.

The need to consider values, methods and use together is further illustrated by the experience of the United Nations Industrial Development Organization (UNIDO). Specifically Margareta de Goys’ article shows that the purpose of evaluation should drive the choice of methods – not the other way round. While striving to improve the relevance, rigour and learning benefits of its evaluations UNIDO discovered the over-

whelming role of enabling contexts in determining outcomes.

As a result, UNIDO evaluators were drawn to systems thinking and contribution analysis to tackle the complexity and uncertainty associated with real life development interventions. Similarly Jan Van Ongevalle's contribution to this issue of *Connections* establishes the limited relevance of results-based planning, monitoring and evaluation approaches compliant with linearity, predictability and control assumptions when dealing with processes of complex change.

In such operating contexts, a mix of actor focused methods combining outcome mapping,

most significant change methods, client satisfaction tools and participatory monitoring works best. In the same vein purposeful monitoring and formative evaluation was Mark Matthews and Geoff White's quest when faced with a request by a State Government in Australia to review its approach to major investments in science and innovation.

In this particular case the *structured hypothesis testing* techniques pioneered by US security agencies proved useful. This involved formulating and testing succinct propositions against summaries of available evidence in a structured and sequential manner. It yielded theories of change better adapted to the task than the audit approach previ-

ously used. Here as elsewhere sharing evaluation knowledge across sectors can generate rich dividends.

In sum this issue of *Connections* confirms that in the real evaluation world values, methods and use are inextricably linked. Accordingly *theory weavers* do deserve special mention in Europe's evaluation space as suggested by Nicoletta Stame. They are the consensus builders of the evaluation enterprise. They help to improve evaluation quality by guiding practitioners through the pathways that link distinct theoretical models at the intersection of knowledge and practice. Your Society's Newsletter serves the very same objective.

## A EUROPEAN EVALUATION THEORY TREE

Nicoletta Stame

This article is based on the deliberations of a Helsinki conference panel about European contributions to the theory of evaluation. Following a summary of my chapter on the "European evaluation theory tree" included in Alkin and Christie's new edition of *Evaluation Roots* (Sage, 2012) three discussants and session participants offered comments.

### The theory tree

Alkin and Christie took the bold step of picturing the vast panoply of theories of evaluation (be they prescriptive, descriptive or explanatory) as a metaphorical tree. First, they labeled its twin sets of roots as "accountability and control" and "social inquiry". Next, they grouped its branches in three bundles: "values", "method" and "use" and used these categories to classify evaluation theories.

Thus they assumed that each evaluation pioneer could be "definitely associated with a particular theoretical position". While they might have dealt with all three theoretical domains they were presumed to have invariably privileged one of them. Hence theorists

were placed on different branches of the tree and asked to talk about the influences that shaped their work, their experience and their perspectives. They were also asked how they felt about their location on the tree.

The first edition of the book was limited to US theorists, and understandingly so since evaluation was mainly "made in the USA" for a very long time. But things have changed and Alkin and Christie decided to open an international window. Patricia Rogers was commissioned to write a chapter on Australia and New Zealand and I was asked to write a chapter about Europe. Since then Carden and Alkin reviewed the work of developing countries theorists.

### The European tree

I found the task challenging: we in Europe either tend to think of ourselves as followers of streams that have sprung elsewhere or are more concerned with doctrines espoused by Europe based institutions and organizations (the European Commission, research centers, etc.) than with individual thought leaders. But after a close search, in-

cluding forgotten papers, I realized that yes, we do have "original thinkers" that match Alkin's and Christie's characterization. We even have a special category of thinkers: the "theory-weavers". These are individuals who have helped to shape an enabling environment for the development of evaluation theories by creating bridges between different approaches and theoretical traditions.

My chapter discovered the following theories (and theorists): illuminative evaluation (Parlett and Hamilton), democratic evaluation (MacDonald); personalized evaluation (Kushner); policy tools (Vedung); dialogue in evaluation (Karlsson); realist evaluation (Pawson and Tilley), syntheses and evidence based policy (Pettigrew, Oakley).

Theory-weavers include Elliot Stern, with his editorship of "Evaluation, the international journal of theory and practice" and his coordination of the EVALSED Guide as well as Eric Monnier (co-author with Jacques Toulemonde of the MEANS guide).

I identified three main ways in which European theorists developed their ideas.

First, I found connections between European theorists and American evaluation thinkers. Here I singled out the original contribution of “democratic evaluation” by such authors as MacDonald, Hamilton and Kushner, who carried out intense dialogues with US authors (e.g. House and Stake).

Second, I uncovered a close link between some European evaluation theorists and Europe’s institutional context. This is especially evident in the Scandinavian group of evaluators who have devised ways of linking evaluation to the democratic tenets of the welfare state. Noteworthy is the work of Vedung on the policy tools. A similar concern prevails among those who originally contributed to creating a European environment more favorable to evaluation practice (the Guides).

Third, some European evaluation thought leaders paid great deal of attention to the evolution of national social policies and European programs. This created an awareness of the complexity of programs and the need for suitable approaches. Thus realist evaluation (Pawson and Tilley) constitutes a distinct and original European approach that focuses on contexts and mechanisms.

## The debate

The panel took up different issues ranging from the general (the overall approach, the characteristics of Europe’s evaluation framework) to the specific (particular approaches).

First, the concept of “theorists” was challenged since it promoted individualism whereas evaluation is a collective practice, e.g. within research institutions. In particular, Helen Simons thought that theories should

be understood within “a wider discourse and institutional practices”. She referred to her own experience with “democratic evaluation” within the East Anglia University group. Its innovations were grounded in the inadequacy of previous methodologies to address controversial value issues. It laid stress on pluralistic ethical principles and inclusive procedures. It adopted a political stance rooted in the need to redress power imbalances when expressing evaluation judgments.

Peter Dahlen Larsen challenged the very idea of “original” thinkers which he wittily described as intent on devising “strategies for eliminating rival authors”. Could it be that by looking for such rare individuals one forgets theoretical connections, over-estimates individuality over dialogue and ignores the shaping of ideas through disagreement? Is it possible to conceive of a “European” evaluation community where different communities co-exist, with different languages and traditions?

Burt Perrin emphasized the distinct European character that contrasts with US attitudes: an appreciation for history and context; a tolerance of cultural differences; a reflexivity and openness to exchanges with other parts of the world. This can go as far as sheer imitation even when flaws have been identified (e.g. the EU obsession with performance indicators). On the other hand, the European sensibility offers the opportunity to address big questions such as those of austerity vs. growth (lest evaluation becomes part of the problem).

Audience reactions had a different flavor. Why include some authors and not others? What empirical justification existed for

the selection of theorists? Was the choice driven by impact (the most read, cited authors)? Or was it motivated by their influence within theory-groups and networks? Finally some participants lamented the scarce representation of their fellow country-people (e.g. Germany).

My response to these observations reflected the original intention of my chapter: to highlight the original contributions of European thinkers. Such contributions are not always recognized, even within European circles. I did not intend to provide a history of European evaluation (that would have included a more representative sample, nationally and by policy sector) nor an account of influential streams of thought, whether commendable or not. But eliciting strong reactions after all, is the risk implicit in any exercise reliant on “connoisseurship” and discretionary judgment.

I hope that entering the game of the evaluation theory tree had the merit of surfacing issues that had been lying underground for a long time:

- How can Europe develop its own theoretical tradition in evaluation?
- What links across the Atlantic and beyond should be nurtured?
- Is there a European evaluation culture or are there different national evaluation cultures within Europe?
- How has evaluation adapted to different cultures?
- How does theory interact with practice and how do ideas develop?
- What is the role of mediators and bridge builders, as embodied by the theory weavers?

## CREDIBLE EVIDENCE – THE CHALLENGES OF MEASUREMENT

Ann M. Doucette, PhD

Evaluation efforts seek to gather information, credible evidence on which to make decisions regarding the effectiveness, efficiency, worth and value of an object of interest (e.g., program, policy, technology, etc.). While we debate the primacy of randomized designs, the sufficiency of quasi-experimental and non-experimental approaches, the clarity of evaluation objectives, the choice among data collection methodologies, the dilemma of data constraints and our differential ability to address the counterfactual; seldom do we scrutinize the actual measures (survey/questionnaire, interview/focus group protocols, etc.) used to gather evidence in evaluation. We applaud the use of sophisticated analytic models that allow us to parcel out the variance attributed to programs and the contribution made by specific components to outcomes, but rarely do we question the soundness of the measures used to support the conclusions made about how or why a program works, how beneficiaries or organizations change, and how effective or efficient interventions are in terms of outcome and impact. More often than not, we assume measurement precision as opposed to scrutinizing the quality of the measurement we rely on to support the conclusions that are reached.

Much of what we measure is not directly or readily observable. Items and questions asked *representationally* define what we cannot directly observe – latent constructs such as distress, equality, corruption, achievement, resilience, and empowerment. We assumed that each item/question provides relevant information and is related to the underlying latent trait. We differentiate respondents in terms of the strength of agreement, reported frequency of an event, and so forth across survey items and questions asked. Responses are summed (or averaged) to parsimoniously assess program progress, effectiveness, efficiency, and value; and, to differentially array outcomes for beneficiaries along a continuum, a “measurement ruler” so to speak, ranging from

favorable to unfavorable results. We use this “ruler” to convey such things as the average amount and range of change experienced as a result of program intervention, supporting our assertions of effectiveness, efficiency, merit and value.

Evaluation design considerations have been intensely debated within the evaluation community (not addressed in this commentary; see Banerjee & Deaton, 2012; Bonell, 2012; Evaltalk Listserv). Conversely, measurement issues and consideration have been largely ignored, other than to note the importance of reliability and validity. An emphasis on reliability and assertions of validity are essential, but insufficient in ensuring the integrity of the evidence yielded from measures.

When we sum or average scores, we assume that all items are equal, and that similar responses across items and questions represent equivalent levels of the underlying latent construct. But is this really the case? In a measure of psychological distress, is an item about sadness equivalent to an item assessing suicidal intent? In assessing multidimensional poverty, are education, health and standard of living equivalent; is deprivation regarding sanitation equivalent to the use of dung, wood or charcoal cooking fuel? In mathematical achievement, is a correct answer to a calculus problem comparable to a correct answer to a more basic long division problem? In simply summing or averaging across responses we assert that all items and questions contribute equally to characterizing the position of the individual, collective, or policy on the measurement ruler. The **content** of questionnaire items and narrative questions asked matters; items do in fact, contribute differentially in positioning individuals and collectives on a measurement ruler (see Figure 1). For example, a response of **agree**, or narrative indication of suicidal ideation more accurately positions an individual as experiencing seriously distress, than does a **strongly agree** response to an item/question about *feeling blue*.

We also assume that the questionnaires and narrative protocols we use in evaluation have sufficient *bandwidth* – measurement coverage of the latent trait(s) of interest. All too often we fail to examine whether the items/questions are sufficient to differentiate respondents, collectives or organizations from one end of the latent construct to the other end. We settle instead for an acceptable reliability as opposed to examining the range of knowledge that can be gained across the items/questions used. Items/questions vary in terms of the level of the trait assessed (content) and may not adequately cover the construct of interest; instead clustering at certain positions along the latent trait continuum. Item/question coverage at the extreme ends of most continuums is typically scant. Inadequate and uneven bandwidth compromises our ability to stably position beneficiaries, collectives, and organizations along the measurement ruler. A few items/questions coupled with a small number of respondents at the ends of the continuum yield unstable estimates of change. Unrecognized item/question gaps result in potential data misinterpretation. For example, stable scores indicating no improvement might well be a consequence of a lack of appropriate items or questions that allow respondents to demonstrate continued improvement or further deterioration (see Figure 1).

Filling in measurement gaps is challenging. In many instances the latent constructs we evaluate are “pseudo-continuums,” at one end of the continuum variation is of interest and is meaningful to the evaluation effort; at the other end it is less so. Although it is challenging to address measurement gaps, knowing that they exist provides more precision in our interpretation of the data, giving us the ability to investigate disappointing outcomes as potential measurement shortcomings (Doucette & Wolf, 2009).

More sophisticated measurement models exist, allowing us to examine the adequacy of item and question coverage – *item response*

theory models (IRT – Lord, 1980; Embretson & Reise, 2000). These models, while prevalent in educational assessment, and medical and clinical investigations, have been substantively slower in terms of their acceptance in evaluation practice. In summary, while measurement is only one step in advancing the credibility of evidence gathered in evaluation efforts; it is nonetheless, its foundation. To ignore the implications of measurement is to conceptualize evaluation as an effort that is subject to the vagaries of measurement artifacts, a substantive compromise to the credibility of the evidence produced.

## Contact Information

Director, The Evaluators' Institute  
 Director, Midge Smith Center  
 for Evaluation Effectiveness  
 Research Professor, Evaluation  
 and Health Policy  
 The George Washington University  
 2147 F Street NW, Suite B-01  
 Washington, DC 20052  
 Phone: 202.994.8112 / eFax: 202.315.3304  
 E-mail: doucette@gwu.edu

## References

Abhijit Banerjee and Angus Deaton, development economists debate the merits of RCTs and alternatives – NYU Development Research Institute, *Debates in Development Conference*. (2012). <http://www.nyudri.org/>

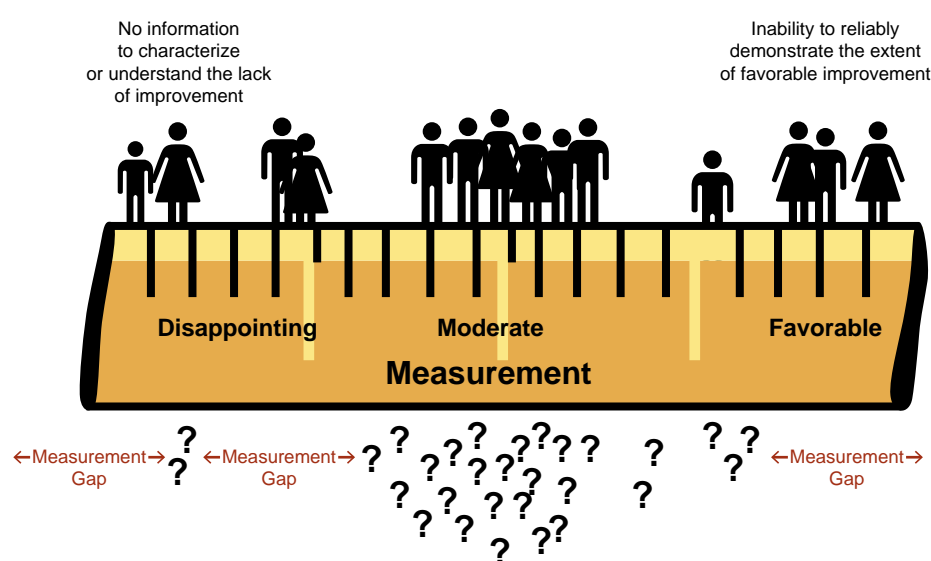


Figure 1: Do measures adequately characterize beneficiaries or programs of interest?

initiatives/deaton-v-banerjee/ – last access 1 August 2013.

Bonell, C.P., Hargreaves J., Cousins, S. et al. (2011). Alternatives to randomization in the evaluation of public health interventions: design challenges and solutions. *Journal Epidemiology and Community Health*, 65(7), 582-587

Doucette, A. & Wolf, A. W. (2009). Questioning the precision of measures used in psychotherapy research. *Psychotherapy Research*, 19(4-5), 374-389.

Embretson, S. E. & Reise, S. P. (2000) *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Evaltalk Listserv Archives – <https://listserv.ua.edu/archives/evaltalk.html> – last access 1 August 2013.

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

Michell, J. (1990), *An Introduction to the Logic of Psychological Measurement*, Hillsdale: Erlbaum.

## RURAL DIFFERENTIATION AND TARGETING: EVALUATING IFAD'S APPROACH<sup>1</sup>

Mattia Prayer Galletti

In 2009, an external peer review of the effectiveness of the Independent Office of Evaluation (IOE) of the International Fund for Agricultural Development (IFAD) recommended further strengthening of evaluation use geared to organizational learning. Amongst other things, IOE was urged to contribute more actively to IFAD's knowledge management activities through the production of

evaluation syntheses and wider dissemination of evaluation findings.

In this context, IOE's synthesis reports emerged as privileged vehicles for the transfer of relevant evaluation experience and lessons learned. For example, IOE's synthesis report on Rural Differentiation and Smallholder Development was prepared with

the intention to contribute to the policy debate triggered by the preparation of the last IFAD's Strategic Framework 2011–2015.

The Evaluation Synthesis was based on a comprehensive desk review about IFAD's experience on targeting. It covered twenty seven project evaluations, nineteen Country Programme Evaluations (CPEs), six Corporate-



Level Evaluations (CLEs) and six Annual Reports on Results and Impacts (ARRIs). Further, the Synthesis reviewed a sample of country strategies and project designs recently approved (2009–2012) taking account of their geographical distribution among five regions.

The review found that almost two thirds of the projects did not identify specific target groups or capture their diversity and specificity. General terms such as poor, poorer or poorest were often used to identify different groups. Of the balance that did differentiate the target population targeting was often done on the basis of demographic (gender, youth, and indigenous people) rather than socio-economic criteria (income, assets).

Evaluations frequently recognized the weakness in project targeting and highlighted the importance of more clearly identifying beneficiary groups at the project design stage. This recommendation was found in forty seven per cent of project evaluations and 60 per cent of country programme evaluations (CPEs). Twenty seven per cent of project evaluation reports provided further detail by recommending a sharper focus on the poorer segments of the rural population. The underlying premise was that an unclear definition of the target group reduces the likelihood of reaching the intended beneficiaries (i.e. the poorer households).

There is evidence that targeting contributes to better results. A review of the performance ratings disclosed that projects featuring a differentiated targeting achieved a higher average score (4.5)<sup>2</sup> compared to those that did not differentiate (4.0). This could be interpreted as the result of the people-focused and participatory approach during implementation that targeting induces. Coupled with flexible and effective project management, targeting seemed to have facilitated adaptation to the needs of vulnerable groups even if such needs had not been identified precisely during project design.

The decision about whom to target is not simply a design issue. It also requires consensus building during implementation. Thus involving local communities in the identification of poor households, through Participatory Rural Appraisal activities (wealth ranking) or other methods, has proven effective in the selection of the most appropriate poverty criteria and in ensuring local ownership about which households to prioritize during project implementation.

As an example, in Cambodia, the Community-Based Rural Development Project in Kampong Thom and Kampot introduced a new targeting approach at Mid-Term Review (MTR) focusing explicitly on the identification of most vulnerable families (MVs) with the support from commune councillors and village representatives. MVs were then provided with special identity cards to ensure them free access to government services and donor support activities.

Closely related to the choice of 'who to work with' is the decision about how to support them. The importance of devising appropriate development strategies and activities to meet the expressed needs of target groups is a common thread in evaluation reports' recommendations. There is often a "disconnect" between project intent and the ability of target groups to take advantage of project services. Hence the full benefits of targeting are not tapped in the absence of actions supporting capacity building and the empowerment of disadvantaged groups.

The recent IOE's Annual Reports on Results and Impact of IFAD Operations have raised concerns that poorer and vulnerable groups might have benefited less than wealthier groups from project interventions. In view of these findings, it appears that IFAD should further tailor its strategies to enable the poorest to benefit from its interventions without falling into a 'one-size-fits-all' approach that strives to reach different groups with similar interventions.

On the other hand, the review also shows positive results when project designs address clearly identified social differences. For example in the case of indigenous peoples cultural differences proved to be an advantage: it induced adaptation of project designs to the local context. A small number of evaluations of rural youth support projects confirmed this finding. Clearly socio-economic differences while they are hard to identify and monitor do matter to project outcomes.

While project interventions usually align with the expressed needs of the target groups, effective poverty reduction strategies may require actions targeted to other social groups. In some contexts migration and non-farm employment are the only opportunities left for improving the livelihoods of poor rural families. In such cases however IFAD can also help address the impact on agriculture production and community resilience for those left-behind.

Beyond targeting, another important element is the need for effective monitoring that allows identifying issues and taking corrective actions in a timely manner. This is critical particularly in rapidly changing environments, e.g. in post-conflict scenarios or areas affected by rapid economic transformation. Unexpected events may affect specific target groups in ways that could not have been identified in design. Ensuring that project management teams practice due diligence with regard to targeting during implementation is therefore an important part of the development effectiveness equation.

As a closing remark, a well-designed project targeting strategy may increase the likelihood that the focus on the poor and marginalized is not lost during implementation. On the other hand, while pro-poor criteria and targeting are essential they are not sufficient on their own to optimise poverty reduction potential.

<sup>1</sup> This article does not necessarily reflect the views of IFAD management or its independent evaluation department

<sup>2</sup> IFAD uses a rating scale of 6 ranging from highly unsatisfactory to highly satisfactory.

# DOES IMPACT EVALUATION IN DEVELOPMENT MATTER? WELL, IT DEPENDS WHAT IT'S FOR!

Irene Guijtand, Chris Roche

## 1. Introduction

It's never fun to waste one's time. So an impact evaluation (IE) that makes a difference would appear to be a good starting point, right? Yet the utility of IE is not always clear or even questioned. Many of the debates about IE focus on formal definitions of impact or on which method is the right one. These tussles are embedded in a growing IE industry with vast amounts of money and many professional reputations at stake. Re-orienting the debate towards the core functions and purposes of IE would help defuse the tedious and costly method-wars that IE has triggered in the development arena.

## 2. So what is the problem?

Our experiences suggest that IE can contribute meaningfully to three purposes.

**Learning to 'improve' as well as 'prove'** what works, or what does not. Many organisations struggle daily with balancing the need to show what strategy is working (i.e. prove) with identifying how to improve implementation (i.e. improve). Both are necessary and linked but undeniably exist in tension. Making IE learning-oriented means considering not only *whose* views and perspectives are central for understanding and valuing impacts in different contexts, but also *whose* learning counts.

**Accountability for resources used.** IEs are often based on a notion of accountability as a contractual relationship between donors and implementers. But fair contracts are difficult to design when precise objectives, or the means to achieve those objectives, are not specified in advance (Shutt 2012). The dominance of *contractual* considerations can also block constructive thinking about alternative forms of accountability that IEs could and should support (Booth 2012). Social or mutual accountability can open up space for IEs of interest to many. Political ac-

countability would help shape IE processes based on significant inputs from citizens that reflect their preferences and interests.

**Influencing for empowerment.** Besides influencing the practice of agencies, IE studies have the potential for much wider use by holding all partners to account. The findings should at the very least be not only accessible but also understandable to those who might want to scrutinise and contest the findings, or use them to promote better policy and practice. We need to invest much more in ensuring that findings can be used by citizens whose welfare needs drive international development, at least in theory.

## 3. So what matters?

A constructive debate about IE transcends disciplinary squabbles and methodological tussles about which method or definition is best. We offer five elements as a starting point for clarifying assumptions and broadening the scope of IE, each of which has practical implications.

**1. Standards matter.** No one contests the importance of having standards by which an intervention or change process can be judged as being 'effective' or norms that the IE methodology should respect. However we argue that greater care is needed about drawing inferences from individual studies or constellation of studies. Effective IE requires looking more holistically at standards not just in IE design but in implementation *and* use.

**2. Rigour and utility matter.** Rigour is one of the most contested aspects of quality within IE often narrowly defined as statistical rigour and ascertaining causality. We need 'relevant rigour' – a standard for rigour that is relevant to the purposes of IE. We also need to be rigorous about ensuring the IE is relevant – 'rigorous relevance'. We also need to be rigorous about assessing which domains of a given

intervention require knowledge and, therefore, merit privileged attention. Rigour and relevance are intricately linked and determine the extent to which an IE is fit for purpose and methodologically appropriate to the nature of the intervention and the context.

**3. Power and politics matter.** Power relations affect evaluative processes in many different ways: in the choice of what is evaluated, how it is done, who is involved, what is valued by whom, its eventual use – and above all who decides all of the above. The politics of IE are more intense when findings are likely to affect people's status, position and livelihoods. Factoring power and politics into IE is important no matter what method is selected – we ignore it at our peril.

**4. Evidence matters.** Evidence clearly matters. No one is against evidence. However we see the need for different types and sources of evidence – depending on the question, the type of intervention and the purpose of a given evaluation. IE largely ignores two important forms of data. First, evidence that tracks the *emerging* impact of programmes to enable ongoing feedback and adapting implementation. Second is the recent emphasis on 'big data', including direct and focused citizen feedback encouraged by organisations like Ushahidi and Twaweza, and real-time data collection from multiple sources through for example the UN's Global Pulse initiative (Kirkpatrick, 2013).

**5. Uncertainty and complexity matter.** Not everyone has signed up to the notion that development work is largely focused on addressing 'wicked problems'. However, a consensus is emerging that development involves systemic change which, by definition, is neither linear nor predictable (Beinhocker 2006, Ramalingam & Jones 2008). Hence, in many contexts IEs need to go beyond proving causality to understanding patterns and correlations that can reduce uncertainty.



## Conclusion

Impact evaluation can contribute to international development effectiveness by improving learning, accountability and by influencing policy and practice. If we agree that these purposes are important then we need to be clearer about *whose* learning counts, *whose* accountability is improved and *whose* influence is strengthened. Furthermore we suggest five topics that can help clarify *how* IE contributes to these processes: standards; rigour and relevance; power and politics; the nature of evidence; and complexity and uncertainty.

For this debate to be even more productive requires a genuine desire to navigate with curiosity and listen across disciplinary codes. We all need to be open to the possibility that our preferred approaches to IE are as likely

to have limitations in terms of the purposes described above – just as the projects that we evaluate.

## References

Beinhocker, E. 2007. *The Origin of Wealth: Evolution, Complexity, And the Radical Remaking of Economics*. Random House Business.

Booth, D. 2012 'Development as a Collective Action Problem', Africa Power and Politics Programme Policy Brief 9,D

Brown, D. L. and Jagadanda. 2007. *Civil Society Legitimacy and Accountability: Issues and Challenges*. Hauser Centre and Civicus. [http://www.civicus.org/new/media/LTA\\_ScopingPaper.pdf](http://www.civicus.org/new/media/LTA_ScopingPaper.pdf)

Guijt, I. 2008. *Seeking Surprise. Rethinking monitoring for collective learning in rural resource management*. Published PhD thesis. Wageningen University and Research Centre.

Kirkpatrick, R. 2013, *Evaluation for the 21<sup>st</sup> Century*, Connections, March

Ramalingam, B and H. Jones, with T. Reba and J. Young. 2008. *Exploring the science of complexity: Ideas and implications for development and humanitarian efforts*. Overseas Development Institute. <http://www.odl.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/833.pdf>

Shutt, C. 2012. *Transformational development and the politics of evidence?* <http://bigpushforward.net/archives/1688>

## LET THE LEARNING DEFINE THE METHOD RATHER THAN THE METHOD DEFINING THE LEARNING

Margareta de Goys

Poverty reduction is an overarching development objective for the United Nations Industrial Development Organization (UNIDO). Many UNIDO evaluations have attempted to assess the impact of UNIDO's interventions on poverty reduction. This met with various challenges but also resulted in interesting findings about results and impact drivers.

Typically a multidimensional definition of poverty was used along five interrelated dimensions (economic, protective, political, socio-cultural and human) influenced by two cross-cutting themes (gender and environment). Evaluations have confirmed that many UNIDO projects contribute to poverty reduction or hold the potential for doing so especially with respect to the first Millennium Development Goal (Eradicate extreme poverty and hunger). Such contributions may be intended or unintended – direct as well as indirect. Most often they are found

within the economic dimension of poverty through job creation or increased income but there were notable examples of other contributions to human dimension. Evaluations also indicate that many opportunities for adopting pro-poor approaches have not been tapped.

The evaluations used a theory based approach, identified impact drivers and included recommendations to improve pro-poor targeting. Findings disclosed long and complex intervention logics involving many intermediate stages between project outputs and ultimate impact. The identification of impact drivers and the adoption of realistic assumptions were important success factors. This was especially the case for policy-oriented and institution building projects.

An encouraging move towards rigorous impact evaluations is underway. However, quantitative evaluations with a control group

are not always possible and at times such approaches do not generate the information we need. Ideally impact evaluations need to be planned for and control groups established at the project design stage. This is not always feasible since as many UNIDO projects focus on supporting the establishment or the strengthening of public sector institutions, such as laboratories, that form part of a national quality system. For ethical as well as legal reasons governments cannot prevent access to public services over a longer period of time and the establishment of control groups are not always possible. Despite these constraints, attempts have been made to assess the impact of UNIDO projects, bearing in mind the poverty reduction, trade capacity building and green industry mandates of the organization. Most impact evaluations used a theory-based approach. Some adopted a systems approach and in a few cases we experimented with systems dynamics modeling in complex

project environments where influencing and contributing factors were many and could not easily be captured in a linear model.

This latter approach demonstrated the importance of looking at not only the project but also the context in which it operates. We drew inspiration from the work of evaluations scholars such as Bob Williams and Richard Hummelbrunner (see June 2013 issue of *Connections*). This approach was challenging but provided opportunities for systemic thinking/analysis and helped to raise awareness of the potential of systems thinking within UNIDO and to ascertain its utility for project design. We were also able to verify that as John Mayne and others have pointed out “projects” are only one of several entities/actions that *contribute* to results.

Systems analysis allows for loop learning and reflection on what causes a change that will, in its turn, cause other changes. It enables contexts or external factors to be part of the analysis thus reaching out beyond the project boundaries and allowing for multiple perspectives (beyond that of the project protagonists) and holistic analyses. For us, the main purpose was to understand if a project concept works and most importantly why.

Our evaluations demonstrated that many influencing factors drive (or fail to drive) impact. We learnt that impact evaluation/analysis should not limit itself to factors within the control of a project. It also needs to look at influencing factors in order to explain the impacts to which the project likely contributed. The impact assessed was in terms of employment, export, income or

consumer protection against sub-standard products.

In a nutshell, we found that:

- the same intervention (defined through the intervention logic or theory of change) can produce different effects not only because the implementation is more or less effective or efficient but because the context is different
- we cannot explain impact purely by our own interventions
- we can generate information pointing to plausible contributions

As an example, a trade capacity building project can reach its objectives in terms of increased capacities to trade and economic development but this does not automatically generate poverty reduction effects for workers in the supported sector. Improved working conditions or increased salaries (for workers receiving salaries below the poverty line) often depend on external (beyond the project) contexts such as consumer awareness and export market regulations.

Of course the United Nations has an important role in the promotion of Corporate Social Responsibility and in raising awareness about social and environmental issues. Evaluation findings clearly pointed to the importance of combining the technical cooperation function with the normative or advocacy roles.

Furthermore, evaluations showed that many so called impact drivers were not known or accounted for at the project formulation stage. Actual contributions were often unintended and could be both direct and

indirect. It is thus not possible to capture and learn about these drivers by purely using a theory-based or control group approach. Our evaluative experience also confirmed the importance of assessing results beyond the outcome level as positive results at one level can be offset by negative effects at another.

For instance positive effects in terms of employment or income can be offset by growing inequalities or negative effects caused by working in a hazardous environment or to the environment. In fact, many gender, socio-cultural, environmental or human dimensions of poverty need to be specifically targeted and negative socio-cultural or environmental effects that might occur need to be monitored and addressed, as appropriate.

The data generated from various methods and subsequent analyses have clearly contributed to an improved understanding on why and how development effects are generated and which are the important impact drivers but also on how the context or the system the project is operating in influence project specific results. This has contributed to organizational learning and helped shape future UNIDO programmes.

In conclusion, effects of development interventions are direct and indirect, foreseen and not foreseen and the quest to understand and measure these effects and to develop models to capture them and to learn why they occur or not needs to continue both for increased learning and for aid effectiveness. For more information about UNIDO evaluations please visit [www.unido.org](http://www.unido.org).

## DEALING WITH COMPLEXITY THROUGH “ACTOR-FOCUSED” PLANNING, MONITORING & EVALUATION (PME)

Jan Van Ongevalle

A growing body of literature points towards the limited relevance of results-based PME approaches that follow a logic of linearity, predictability and control when dealing with processes of complex change characterised by emergence, unpredictability and non-linearity (Forss et al., 2011). In addition, programmes supporting complex change are often built around actors who hold different understandings of the programme's objectives, how to achieve these, and what roles and responsibilities they should assume (Jones, 2011). This leads to change processes that are much more difficult to predict, measure and understand (Stern et al., 2012).

### An assessment framework for complexity oriented PME

To assess to what extent a PME approach is complexity oriented, an analytical framework was developed by a three-year action research project (2010–2012) in which ten development organisations (nine Dutch and one Belgian) together with their Southern partners and four research coordinators explored how a variety of PME approaches help respond to a results agenda while dealing with processes of complex change. This framework guided the action research. It consists of four questions relevant to four implications of complexity for PME practice previously identified through a review of literature on PME and complexity:

1. To what extent has the PME approach helped programme stakeholders to:
  - clarify their expectations
  - clarify their roles and responsibilities
  - strengthen trustful relationships
2. To what extent has the PME approach helped programme stakeholders to:
  - learn about the effects of the programme
  - track effects that are difficult to quantify
  - learn about unexpected effects
  - strengthen processes of collaborative learning

3. To what extent has the PME approach helped programme stakeholders to:
  - satisfy upward accountability needs of the donor
  - satisfy horizontal accountability needs
  - satisfy downward accountability needs of the beneficiaries
4. To what extent has the PME approach helped programme stakeholders to:
  - contribute to changes in the internal practices of the programme
  - gain clarity about the programme's contribution to the observed effects.

### Actor-focused PME

The PME approaches piloted through action research included outcome mapping (OM), most significant change (MSC), client satisfaction tools (CSI), Sensemaker and participatory M&E tools such as person goal exercises. A common characteristic that made these approaches attractive turned out to be their focus on the actors that the programme was trying to influence directly or indirectly (see figure).

Rather than focusing on the ‘hoped for changes in state’ (e.g. changes in income levels) they direct attention to what people do (e.g. behaviour, practices, relationships) in order to contribute to the hoped-for changes in state and/or people's perceptions. This contributed to advantages as well as challenges:

1. It provided a language for deeper conversations among programme stakeholders and helped strengthen relationships. The increased social interaction was also shown to contribute to a shared actor focused theory of change. More insight was gained through monitoring. An important challenge was the need for strong leadership to sustain actor focused PME practice. This involved ‘unlearning’ some conventional ways of doing PME.
2. The particular focus on changes in behaviour, professional practice, relationships and perceptions helped learn about

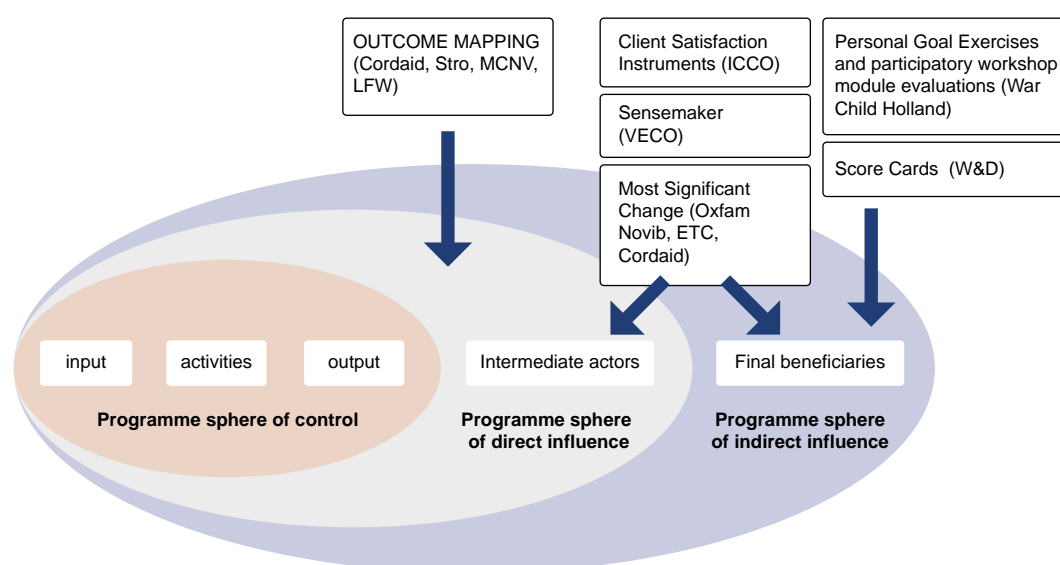
results, that would otherwise be missed. However, deeper analysis and collective sense making remained a challenge requiring customized facilitation.

3. The ability to report about a wide variety of results (i.e. changes in behaviour and relationships) contributed to improved accountability. In addition, the increased social interaction was seen to enhance transparency, trust and empowerment of stakeholders. Furthermore, issues of bias or ‘cherry picking’ the positive stories became less of an issue since monitoring information was firstly used for critical reflection about the programme and allowed different stakeholders with different perspectives to contribute their inputs. However, using the feedback for learning and providing the necessary time and resources to inform the people who provided it about what was done with it remained a considerable challenge.
4. Actor focused PME approaches were shown to promote wider involvement of staff in reflection and learning. It directed field staff and partners to focus on the effects of a programme instead of the limited preoccupation with programme processes and activities. However, too much dependence on external consultants increased the risk of “externalising” the learning process. Furthermore, there were instances where lessons learned through actor centred PME were ignored by management because they didn't recognise them as valid outcomes of mainstream PME practice.

### Recommendations for PME practitioners

- Actor focused PME approaches can help to broaden a programme's “radar screen”. In other words, tracking changes in behaviour, relationships or practices of target groups in different spheres of influence can bring to light important information about a programme's effects or results that would otherwise remain hidden.

Visualisation of actor focus of the PME approaches piloted in the action research.



- Regular monitoring resulting in lessons that inform programme adjustments is key when dealing with complex change processes. This requires shifting perceptions of the meaning and value of regular monitoring practice. Monitoring is still widely perceived as the little brother (or sister) within monitoring and evaluation. It is often limited to a regular follow-up of programme activities and activity-based progress reports. Evaluation, on the other hand, is associated with deeper learning about programme results. At the same time, such evaluations are often carried out by external consultants, and the learning is therefore externalised.
- Monitoring by programme staff needs to go beyond activity monitoring but should systematically and regularly track programme effects and motivate learning processes that can inform programme adjustment whenever the programme is getting off track. Strong leadership that motivates and mandates regular learning-centred monitoring of programme effects is essential.

### Recommendations for donors

- There is need to adopt a wider notion of what results may entail and accept as

valuable results changes in behaviours, relations or perceptions among social actors directly or indirectly influenced by a programme.

- Donors should ask funded programme actors to demonstrate that they have developed and implemented PME systems that are learning-centred and require specific accounts of how lessons learned were used for programme improvement or planning.
- Donors should insist on funding proposals that are clear and explicit about the various actors in a programme's sphere of control (i.e. who is responsible for inputs, activities, outputs), spheres of direct influence (direct target groups) and spheres of indirect influence (indirect target groups or/and final beneficiaries).
- Finally, appreciation should be shown for programmes that demonstrate a deep understanding of theory of change and a readiness to adapt the original theory of change as implementation progresses. This would help safeguard and promote flexibility in programme planning and execution which is essential when supporting complex change processes.

### References

1. Forss, K., Marra, M. and Schwartz, R. (2011). *Evaluating the Complex. Attribution, Contribution and Beyond. Comparative Policy Evaluation, Volume 18*, London, Transaction Publishers.
2. Guijt, I. (2010) Rethinking monitoring in a complex messy partnership in Brazil, *Development in practice*, Vol. 20:8, p.1027-1044.
3. Jones, H. (2011). 'Taking responsibility for complexity. How implementation can achieve results in the face of complex problems. Working Paper 330', London, UK: Overseas Development Institute (ODI). Available at: <http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/6485.pdf>
4. Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., Befani, B. (2012) Broadening the range of designs and methods for impact evaluation. Study Report, DFID, London, UK. Available at <http://mande.co.uk/blog/wp-content/uploads/2012/06/2012-Stern-Working-Paper-38-final-April-20121.pdf>

## FASTER, SMARTER AND CHEAPER: HYPOTHESIS-TESTING IN POLICY AND PROGRAM EVALUATION

Mark Matthews and Geoff White

We were commissioned by a State government in Australia to review its approach to the evaluation of a major investment in science and innovation that it made over a decade ago. It quickly became clear to us that the traditional auditing process adopted by our client which relied on reconstruction of the original objectives and targets of the investment was inappropriate. It boosted costs, slowed down the evaluation and sapped the goodwill of stakeholders – without contributing much by way of evidence on the impact and value for money of the investment.

Apart from the obvious finding that an evaluation framework should have been specified at the outset of the intervention, we concluded that the audit approach was cumbersome and expensive. It required reconsideration. Accordingly, we suggested an alternative designed to increase the speed, effectiveness and efficiency of the evaluation. Our advice might have wider applicability not only in the science and innovation policy field but more generally for the assessment of major investments implemented in uncertain contexts.

The specific solution we proposed drew on the *structured hypothesis testing* techniques used by the US security intelligence community. It involves assessing the investment by formulating and testing succinct propositions against summaries of available evidence in a structured and sequential manner. It reverses the time-line of the audit approach (tracking the intervention as it unfolds over time) by identifying the contribution of the intervention from a formative perspective – how we might move forward taking account of current budgetary and policy priorities.

The approach aims to maximize the signal to noise ratio by quickly testing hypotheses against the balance of probabilities and by reporting the results in a concise fashion ac-

cording to a format suitable for use at later dates. A critical aspect of the approach is that it allows for, and encourages, an iterative process by which hypotheses are readily reformulated and evaluation results revised as fresh evidence becomes available – without recourse to major and costly re-writing of unwieldy reports.

The hypothesis testing approach we piloted proved to be effective in getting the State government evaluation back on track and engaging stakeholders in a positive dialogue about preparing for the future (rather than just reflecting on the past). It also demonstrated potentially significant cost savings compared with the audit-based approach. Consequently, the State government is now adopting the approach more widely and further developing its key components, namely:

- Engagement of stakeholders in jointly specifying and agreeing the key hypotheses underpinning the intervention program to be tested and adjusted as development proceeds
- Agreement amongst stakeholders at the outset of the program on the design of the monitoring and evaluation questions, framework, methods and reporting arrangements and their respective roles in providing the necessary data
- Mixed evaluation methods and a process of triangulation to assess convergence of the emerging evidence – to reduce complexity as the intervention unfolds
- Tracking mechanisms to ensure that the evaluation can help reduce uncertainty and inform decisions in a clear and simple way especially at anticipated ‘forks’ in the road of program development
- Methods by which evaluation design and reporting can be standardized and the integrity of the evaluation process protected – including protocols that guarantee evaluators’ independence regardless of whether they are internally or externally commissioned.

The approach we piloted is promising for more widespread use given its following compelling attributes.

Firstly, by reducing the complexity and workload of evaluations it lowers their cost and duration. If applied in a continuous developmental manner, such evaluations could spot early program failures and contribute to better use of resources.

Secondly, it offers a clear and standardized discipline for the design and conduct of program evaluations. It enables public authorities to become smarter commissioners of evaluations, recipients of government funding to be clearer about evaluation information needs, and evaluators to be more consistent in the conduct and reporting of evaluations.

Finally, the approach can be set up at the start of a program and at low cost. Hence, program managers will know at the outset what the key evaluation issues are and how they will be assessed. The resulting discipline of continuously monitoring and evaluating how public value is being generated (or not) by testing hypotheses against available evidence increases the likelihood of generating benefits for taxpayers by fostering continuous improvement.

On the basis of experience to date, the hypothesis testing method could help to bridge theory of change approaches and real time developmental evaluation in a pragmatic and rigorous way. It could contribute to the integration of policy and program design with the evaluation methods to be used before, during and after the intervention.

This potential exists because policy and program design in complex and uncertain contexts is itself often based on hypotheses about the underlying theory of change and the appropriate form for program architectures and delivery mechanisms. Consequently, the approach can also be used to collate and analyse the evidence that drives policy

and program design by integrating, assessing and reporting research findings from different interventions in a single comprehensive framework.

The next phase in the development of the approach is a set of pilot activities to be carried out in partnership with interested

Australian federal government departments and agencies to explore the potential of this integrated approach (including pilots in analysing complex evidence that informs policy stances). There is the potential for wider experimentation with developing the approach. For example, it might make sense to set up an 'open source' forum – allowing the re-

sults of multiple pilots to be logged, lessons learned to be spelled out and a community of good practice to be developed. Further information on the approach can be obtained from the authors on request.

The corresponding author is Mark Matthews (mark.matthews@anu.edu.au).

## THE AUTHORS

### **María Bustelo**

Maria Bustelo, PhD in Political Science, is Associate Professor in the Department of Political Science and Public Administration at the Complutense University, Madrid (UCM), Spain. She has worked as an independent consultant for Spanish public administrations and agencies, the European Commission, and NGOs in the field of programme evaluation, and in the areas of community development, health promotion, drug dependency prevention, and gender policies. She is also the director of the Master on Evaluation of Programmes and Public Policies (UCM), and member of the Board of Director Committee of the National Agency for the Evaluation of Public Policies and the Quality of Public Services. She leads the European QUING project team at the UCM (QUING: Quality in Gender+ Equality Policies -2006–2011- [www.quing.eu](http://www.quing.eu); a research project supported by the VI European Framework Programme). She has a number of publications on analysis and evaluation of public policies and gender equality policies.



### **Mattia Prayer Galletti**

Mr. Mattia PRAYER GALLETTI, a national of Italy, joined IFAD in 1988. Since 1989 he has been working as Country Programme Manager in the Asia and the Pacific Division, responsible for IFAD-supported programs in Bhutan, Laos, Mongolia, Papua New Guinea, China, Vietnam, Indonesia, Iran, the Pacific region and India. In 2011 he moved as Senior Evaluation Officer in the Independent Office of Evaluation. Mr. Prayer Galletti holds a B.A.



in Political Science from the University of Bologna, and two Masters in Economics of Agro-Food Systems (Cremona, Italy) and in Development Studies (Geneva, Switzerland).

### **Margareta de Goys**

Margareta de Goys is the Director of UNIDO's Evaluation Group. She was the Vice-Chair of the United Nations Evaluation Group (UNEG) in 2011/12. Previously she was a partner and senior consultant of SPM Consultants, Stockholm, Sweden. She has been involved in a large number of bilateral and multilateral evaluations, both in the capacity of evaluation manager and evaluator. She was the Co-Chair of the NONIE 2012 meeting.



### **Ann M. Doucette, Ph.D.,**

is the Director of The Evaluators' Institute (TEI) and the Midge Smith Center for Evaluation Effectiveness, and Research Professor of Evaluation and Health Policy at The George Washington University, USA. Her expertise includes the development of rigorous and credible performance measurement systems that target accountability, quality monitoring, outcomes and impact for system, organization, and individual levels of program intervention. This work includes specialized emphasis on measurement, which Dr. Doucette considers fundamentally critical in performance monitoring, evaluation practice, and outcome/impact assessment. Her work in these areas most often emanates from a complex adaptive systems perspective.



### **Dr. Irene Guijt**

is an independent consultant, part-time Research Associate at the Overseas Development Institute and a Visiting scholar at the Australian National University. She has worked in international development for over 20 years, emphasising participatory processes. She undertakes research, system design, facilitation and advisory work on learning-oriented knowledge processes in international development. Recent work includes piloting SenseMaker for evaluation in East Africa and Asia, a programme of work on impact evaluation with AusAID, and piloting participatory impact assessment for IFAD. She is active in global capacity building through BetterEvaluation and in the politics of evaluation as co-convenor of the Big Push Forward.



### **Mark Matthews**

is the Executive Director of the HC Coombs Policy Forum at the Crawford School of Public Policy, Australian National University (ANU). The Forum is a joint Australian Government – ANU policy think tank that focuses on experimental and exploratory policy work carried out at the interface between government and academia. Mark's own research and policy engagement work focuses on governments' role as uncertainty and risk manager of last resort and the implications of this distinctive role for policy formulation, program evaluation and review, and the specific implications for science and innovation policy (including public sector innovation). In addition to a career as a management consultant, Mark has held academic positions in the UK universities of Sussex, Bath and Warwick.





**Jan Van Ongevalle**

is research manager at the Research Institute for Work and Society (HIVA) at the Catholic University of Leuven where he leads the development cooperation research group (<http://www.hiva.be/en/>). His research interests include complexity oriented monitoring and evaluation, participation and collaborative learning within programmes of international development, global education and awareness raising.

**Robert Picciotto**

Robert ('Bob') Picciotto, (UK) Professor, Kings College (London) was Director General of the World Bank's Independent Evaluation Group from 1992 to 2002. He previously served as Vice President, Corporate Planning and Budgeting and Director, Projects in three of the World Bank's Regions. He currently sits on the United Kingdom Evaluation Society Council and the European Evaluation Society's board. He serves as senior evaluation adviser to the International Fund for Agricultural Development and the Global Environment Fund. He is also a member of the International Advisory Committee on Development Impact which reports to the Secre-



tary of State for International Development of the United Kingdom.

**Chris Roche**

as over 25 years' experience working for International NGOs as a project manager, evaluator, policy researcher and as a Director. He has recently joined La Trobe University as an Associate Professor and is the convenor of the Masters of International Development. He is the author of Impact Assessment for Development Agencies, and co-editor of Ethical Questions and International NGOs. Chris is also engaged in supporting the monitoring and evaluation efforts of the Pacific Leadership Program and the Developmental Leadership Program, for whom he has written two papers on the evaluation of politics and the politics of evaluation.

**Nicoletta Stame**

retired from teaching Social Policy at the University "La Sapienza", Rome. She was a co-founder and the president of the Italian Evaluation Association. She is a Past President of the European Evaluation Society. She is associate editor of *Evaluation*, and participates in the International Evaluation (Inteval) net-



work. Nicoletta is interested in the theory and methods of evaluation. She is author of *L'esperienza della valutazione*, editor of *Classici della valutazione*, co-editor (with Ray Rist) of *From Studies to Streams*, and author of many essays in books and journals. She has evaluated programs of enterprise creation, social integration and R&D at the local, national and European level. Her work aims at enhancing the evaluation capacities of public administrators, program implementers and beneficiaries.

**Geoff White**

is a professional economist with substantial practical policy and consultancy experience in policy evaluation. He held senior posts in the UK Department of Trade and Industry and HM Treasury until 1988 when he became associate director with Coopers & Lybrand and then with PACEC. Geoff joined the SQW consultancy where he directed its UK evaluation practice for fifteen years until late 2010 when he set up his own research and evaluation business. He has carried out evaluations in widely diverse policy areas and contributed to the development of official evaluation guidance (most recently a major review and edit of the Magenta Book – the official UK guide on policy evaluation).

**EES****EUROPEAN  
EVALUATION  
SOCIETY****1–3 October 2014**

Pre-conference workshops 29–30 September 2014

**11<sup>th</sup> EES Biennial  
Conference****Evaluation for an  
Equitable  
Society**[www.ees2014.eu](http://www.ees2014.eu)**Independence, Partnership, Participation**

## EVALUATION NEWS: PUBLIC INVOLVEMENT IN RESEARCH: GETTING THE MEASURE OF SUCCESS

Public involvement in research has grown significantly in recent years. There are now tens of thousands of people working alongside academics and funders designing and conducting research on subjects ranging from health to physics and from local history to conservation.

A Public Involvement Impact Assessment Framework (PiiAF) was launched on September 6 2013 in London. At the launch researchers, research funders and members of the public learnt about PiiAF and got a chance to try out the online version, which was recently posted at <http://piiاف.org.uk/index.php>.

PiiAF was designed by Universities in Lancaster, Liverpool and Exeter. They joined forces with to produce an innovative resource to help evaluators assess the impact of public involvement in research. PiiAF was developed with funding provided by the Medical

Research Council Methodology Research Programme. The public was involved as user investigators on the research team and through the study's Public Advisory Group and National Advisory Network.

Surveys conducted by PiiAF found that close to 90% of respondents considered it important to involve the public in research impact assessments. More than 200 different types of impacts – on research and on the people involved – were identified in a recent review of literature, some 40% of which negative. However, very little of this literature was grounded in rigorous evaluations.

*"The complexity of public involvement in research makes evaluation very challenging – no single assessment method will cover all situations... However, this new resource will help evaluators to identify the issues that could affect the impacts public involvement can have on their*

*research and to develop an approach to assessing these impacts that is tailored to their particular situation,"* said Jennie Popay, Professor of Sociology and Public Health at Lancaster University, who led the team of researchers and who organised the workshop.

The PiiAF framework helps users develop a pathway from their approach to public involvement through to the impacts they want it to have, to identify questions for their evaluation and to decide on the most appropriate methods to use. Prior testing of the PiiAF resource has shown it will also be a useful resource in training for evaluators and members of the public interested in getting involved in research.

Professor Popay and her team would welcome feedback on the PiiAF from Connections' readers. Please address them to: [j.popay@lancaster.ac.uk](mailto:j.popay@lancaster.ac.uk).

### AN INVITATION TO ALL MEMBERS TO PARTICIPATE IN THE 2013 EES ANNUAL GENERAL MEETING (AGM) AND THE EES BOARD ELECTIONS

This is an invitation to all members to participate in the AGM. It is also an invitation to non-members to join or rejoin the EES. The 2013 EES AGM will be held on 13 December 2013, from 15:00 till 16:00 (CET) in Paris, France. This year's AGM is tasked with the election of two new EES Board members for positions becoming vacant as of January 2014. One of these is for the Vice-President/President Elect to be appointed following Claudine Voyadzis' ascendancy to the Presidency.

Voting will be online starting in early December, running until the AGM date. As provided for in the statutes of the Society (Art. 16) and in order to facilitate full participation by members the EES Board will hold the AGM with the help of electronic communications. We will explain how to join the AGM electronically via an email that will be addressed to EES members in good standing seven days prior to the AGM date. We will announce the results at the end of the AGM.

All EES members are cordially invited to participate. For more information about the AGM agenda, the online election process and the candidates please visit the EES website at [www.europeanevaluation.org](http://www.europeanevaluation.org).



### KARIN ATTSTRÖM, EES BOARD MEMBER LEAVING THE POST

Please join me in thanking Karin for her four-year service as Board Member, especially for her contribution to the EES communication strategy and her sharp insights.

María Bustelo, the EES President.